arXiv:q-bio/0402009v3 [q-bio.PE] 29 Apr 2004

The Speed of Adaptation in Large Asexual Populations

Claus O. Wilke*
Digital Life Laboratory 136-93
Caltech, Pasadena CA 91125

Running head: Speed of adaptation

Keywords: clonal interference, virus evolution, optimal mutation rate, Hill-Robertson effect

Corresponding author:
Claus O. Wilke, Digital Life Laboratory 136-93, Caltech, Pasadena CA 91125
Phone: 626 395 2258, Fax: 626 395 2944, email: wilke@caltech.edu

*Present address:
Keck Graduate Institute of Applied Life Sciences, 535 Watson Drive, Claremont, CA 91711
email: claus_wilke@kgi.edu

Abstract: In large asexual populations, beneficial mutations have to compete with each other for fixation. Here, I derive explicit analytic expressions for the rate of substitution and the mean beneficial effect of fixed mutations, under the assumptions that the population size N is large, that the mean effect of new beneficial mutations is smaller than the mean effect of new deleterious mutations, and that new beneficial mutations are exponentially distributed. As N increases, the rate of substitution approaches a constant, which is equal to the mean effect of new beneficial mutations. The mean effect of fixed mutations continues to grow logarithmically with N. The speed of adaptation, measured as the change of log fitness over time, also grows logarithmically with N for moderately large N, and it grows double-logarithmically for extremely large N. Moreover, I derive a simple formula that determines whether at given N beneficial mutations are expected to compete with each other or go to fixation independently. Finally, I verify all results with numerical simulations.

INTRODUCTION

In asexual populations, beneficial mutations that have arisen independently in different organisms cannot recombine and therefore have to compete for fixation. This effect, often referred to as clonal interference (Gerrish and Lenski 1998), leads to a slowdown of adaptation for large population sizes. A similar effect can arise in sexual populations, and is called the Hill-Robertson effect (Hill and Robertson 1966) or the traffic problem (Stephan 1995; Kirby and Stephan 1996). (See also Crow and Kimura 1965; Kimura and Ohta 1971; Barton 1995; Orr 2000; McVean and Charlesworth 2000; Gerrish 2001; Johnson and Barton 2002; Kim and Stephan 2003) Clonal interference has two main consequences: As the population size becomes large, the increase in the rate of adaptation with increasing population size declines, and the beneficial mutations that are fixed convey increasingly larger beneficial effects. A number of recent studies have tried to quantify the rate of adaptation and the distribution of beneficial mutations in various organisms whose predominant mode of replication is asexual, such as *Escherichia coli* (de Visser et al. 1999; Imhof and Schlötterer 2001; Rozen et al. 2002), vesicular stomatitis virus (Miralles et al. 1999; Miralles et al. 2000), and bacteriophages ΦX174 and G4 (Bull et al. 2000; Kichler Holder and Bull 2001).

Early studies of clonal interference date back to Kimura and coworkers (Crow and Kimura 1965; Kimura and Ohta 1971). These authors considered the same effect s for all beneficial mutations. Gerrish and Lenski (1998) were the first to consider a distribution of beneficial effects, but neglected deleterious mutations. The results of Gerrish and Lenski (1998) were later generalized by Orr (2000) to include deleterious mutations. In the works of both Gerrish and Lenski (1998) and Orr (2000), the final results (formulae for the expected substitution rate and for the mean effect of fixed mutations) were given in the form of unwieldy double integrals, which are difficult to interpret.

[However, Gerrish and Lenski (1998) gave explicit expressions for the unrealistic case of uniformly distributed beneficial mutations.] From these integrals, we cannot easily estimate for what parameter settings the interference effect becomes important, and it is unknown how the speed of adaptation behaves for very large N. Moreover, even numerical evaluation of the integrals can be tricky, because the integrand is strongly peaked. Rozen et al. (2002) gave an explicit expression for the distribution of beneficial effects of fixed mutations at large N. However, this expression also does not lead to a simple expression for the mean.

Here, I derive asymptotic expansions for the expected rate of adaptation and for the mean beneficial effect of fixed mutations, under the assumption that beneficial mutations are distributed exponentially. This assumption is reasonable, and has good theoretical support from extreme-value theory (Gillespie 1983; Gillespie 1991; Orr 2003). I find that for very large N, the expected rate of adaptation approaches a limiting value that is given by the mean selective advantage of new beneficial mutations. The mean beneficial effect of fixed mutations, on the other hand, does not reach a hard limit, but continues to grow with the logarithm of the population size.

MATERIALS AND METHODS

Model: I consider the model analyzed by Orr (2000). I assume that N haploid organisms replicate asexually, and accumulate both deleterious and advantageous mutations. The total mutation rate per genome and generation is U, and the fraction of beneficial mutations is p_b. Hence, the beneficial mutation rate is Up_b, and the deleterious mutation rate is $U(1-p_\mathrm{b})$ ($\approx U$ for small p_b). The effects of both beneficial and deleterious mutations are drawn from probability distributions; all mutations act multiplicatively. I

use a slightly simplified notation for beneficial and deleterious effects of mutations in comparison to Orr (2000): By s, I denote the effect of a particular mutation, either beneficial (in which case fitness is increased by a factor $1+s$) or deleterious (in which case fitness is decreased by a factor $1-s$). The mean effect of beneficial mutations is s_b, and the mean effect of deleterious mutations is s_d. The harmonic mean of the distribution of deleterious mutations is s_H. At equilibrium (when all beneficial mutations have gone to fixation), the frequency of the class of individuals with the highest fitness is approximately $P_0 = \exp(-U/s_\mathrm{H})$ (Orr 2000).

I assume that beneficial mutations are exponentially distributed, that is, beneficial effects are drawn from a distribution with probability density function $f(s) = \exp(-s/s_\mathrm{b})/s_\mathrm{b}$. The analytic calculations make no assumption about the distribution of deleterious mutations, but all simulations have been carried out with a truncated exponential distribution (see Simulation methods). I assume that on average deleterious mutations have a much larger effect than beneficial mutations ($s_\mathrm{b} \ll s_\mathrm{d}$), such that beneficial mutations rarely compensate deleterious mutations.

Simulation methods: I carried out simulations of the model described in the previous subsection. In the simulations, N sequences were propagated in discrete generations. The number of offspring sequences of a sequence i in the next generation was binomially distributed with mean $w_i/\langle w\rangle$, where w_i is the fitness of sequence i and $\langle w\rangle$ is the average fitness of the population. Each offspring sequence suffered k_b beneficial and k_d deleterious mutations, where k_b and k_d were Poisson-distributed with means Up_b and $U(1-p_\mathrm{b})$, respectively. Each beneficial mutation increased the fitness of a sequence by a factor of $1+s$, where s was drawn from an exponential distribution with mean s_b. Each deleterious mutation decreased the fitness of a sequence by a

factor of $1-s$, where s was drawn from a truncated exponential distribution with parameter a. The distribution was truncated both to the left and to the right. The left truncation was necessary to avoid a zero harmonic mean s_{H}. (For $s_{\mathrm{H}} = 0$, the predicted frequency of the unmutated individuals is $P_0 = 0$, and the theory breaks down.) I used as a cutoff for the left truncation the value 0.01. The right truncation is necessary to avoid negative fitness, and here I used the cutoff value 1. As parameter for the truncated exponential distribution, I used $a = 0.1$, which results in $s_{\mathrm{d}} = 0.11$ and $s_{\mathrm{H}} = 0.05$.

I let the population equilibrate for 1000 generations at $p_{\mathrm{b}} = 0$ before I set p_{b} to its desired value and started measuring the rate of adaptation. Simulations were continued for up to 50,000 generations, depending on population size (the smaller the population size, the longer the simulation run), and replicated between 5 and 50 times (the smaller the population size, the more replicates). For each sequence in the population, I kept track of the number of beneficial mutations it had accumulated. At the end of a simulation run, I subdivided the final population into classes with equal numbers of beneficial mutations and determined the most abundant class. The number of beneficial mutations n in the most abundant class divided by the number of generations since equilibration Δt served as an estimator for the rate of substitution k. I averaged k over all replicates to obtain the result reported for $\mathrm{E}[k]$. In order to obtain an estimate for the change in log fitness over time $d \log w(t)/dt$, I determined the sequence with the least number of deleterious mutations in the most abundant class, and divided the logarithm of the sequence's fitness by Δt. Again, I averaged over all replicates to arrive at the values reported here. To test whether this approach was comparable to a direct measurement of the change in population fitness, I fitted for several exemplary runs a straight line to the logarithm of the average population fitness

as a function of time, from the end of the equilibration time to the end of the simulation run, and took the slope of that line as the value for $d \log w(t)/dt$. The differences in the results obtained with these two alternative approaches were minute.

RESULTS

Expected substitution rate and mean beneficial effect: Beneficial mutations arise in the populations at rate NUp_b. If this rate is small, then they do not interfere with each other, and independently go to fixation or are lost to drift. In this case, their expected probability of fixation (averaged over all possible beneficial effects) is $2s_\mathrm{b}P_0$ (Orr and Kim 1998; Campos 2003), and therefore the expected rate of substitution E[k] becomes (Orr 2000)

$$\mathrm{E}[k] = 2NUp_\mathrm{b}s_\mathrm{b}P_0\,. \tag{1}$$

When beneficial mutations interfere with each other, then their probability of fixation is reduced by a factor of $e^{-I(s)}$, with $I(s) = 2Up_\mathrm{b}P_0N \ln N(s + s_\mathrm{b})s^{-1}e^{-s/s_\mathrm{b}}$ (Gerrish and Lenski 1998; Orr 2000). $I(s)$ is the expected number of new mutations of effect larger than s that occur in the time interval of length $t = (2/s)\ln N$ during which a mutation of effect s goes to fixation. The form of $I(s)$ that I use throughout this article assumes that beneficial mutations are distributed exponentially. The general form for arbitrary distributions is given in (Gerrish and Lenski 1998; Orr 2000). The expected rate of substitution is obtained by integrating over all beneficial mutations. Again using the assumption that beneficial mutations are exponentially distributed, one finds that (Gerrish and Lenski 1998; Orr 2000)

$$\mathrm{E}[k] = 2NUp_\mathrm{b}P_0s_\mathrm{b}^{-1}\int_0^\infty se^{-I(s)-s/s_b}ds\,. \tag{2}$$

In Appendix 1, I show that for large N, the substitution rate becomes

$$\mathrm{E}[k] \approx \frac{s_\mathrm{b}}{\ln N}[\ln(2Up_\mathrm{b}P_0N\ln N) + 0.5772]\,. \tag{3}$$

In the limit of $N \to \infty$, this expression simplifies to

$$\mathrm{E}[k] \approx s_\mathrm{b}\,, \tag{4}$$

that is, the rate of substitution reaches a hard limit that is given by the mean beneficial effect of new mutations. Figure 1 shows that the approximation Eq. (3) works well for intermediate to large N. However, $\mathrm{E}[k]$ comes close to its limiting value s_b only for very large N.

According to Gerrish and Lenski (1998), we can calculate the mean beneficial effect of fixed mutations $\mathrm{E}[s]$ as

$$\mathrm{E}[s] = \frac{\int_0^\infty s^2 e^{-I(s)-s/s_\mathrm{b}} ds}{\int_0^\infty s e^{-I(s)-s/s_\mathrm{b}} ds}\,. \tag{5}$$

This expression simplifies to (see Appendix 1 for details)

$$\mathrm{E}[s] \approx s_\mathrm{b}[\ln(2Up_\mathrm{b}P_0N\ln N) + 0.5772] \tag{6}$$

for large N. Figure 2 shows that this approximation also works very well for intermediate to large N.

Estimating the onset of clonal interference: For small N, the expected substitution rate is $\mathrm{E}[k] \approx 2NUp_\mathrm{b}P_0s_\mathrm{b}$ [Eq. (1)], while for very large N, we have $\mathrm{E}[k] = s_\mathrm{b}$. On the basis of these two equations, we can derive a simple estimate for the parameter regions in which clonal interference is relevant: We are certainly in the clonal-interference regime if the estimate of $\mathrm{E}[k]$ for small N exceeds that for large N, that is, if

$$N > \frac{1}{2Up_\mathrm{b}P_0}\,. \tag{7}$$

This result has a simple interpretation: Clonal interference becomes relevant if—on average—one beneficial mutation arises in the zero-mutation class at least every other generation. Note that the mean effect of deleterious mutations enters this result (through P_0), but not the mean effect of beneficial mutations.

The estimate Eq. (7) is fairly conservative, in the sense that when N exceeds $1/(2Up_{\rm b}P_0)$, we are sure that clonal interference is important, but clonal interference starts having some effect already for smaller N. In Appendix 1, I show that an improved estimate is

$$N \ln N > \frac{1}{2Up_{\rm b}P_0} \,. \tag{8}$$

Figure 1 illustrates where the two estimates Eqs. (7) and (8) lie with respect to the exact expression and the approximations for $\mathrm{E}[k]$.

Speed of adaptation: The expected substitution rate is in general not an accurate measure for the speed of adaptation, because it disregards the beneficial effect of the fixed mutations. A better measure is the change in fitness (or log fitness, which is more appropriate for a multiplicative model) over time. Clearly, the faster fitness increases, the faster a population adapts to its environment.

As mentioned by Johnson and Barton (2002), the change in log fitness is given by

$$\frac{d\log w(t)}{dt} = \mathrm{E}[k]\log(1+\mathrm{E}[s])\,. \tag{9}$$

Using $\mathrm{E}[k] = s_{\rm b}$ and $\mathrm{E}[s]$ as given in Eq. (6), we find for large N

$$\frac{d\log w(t)}{dt} \approx s_{\rm b}\ln[1+s_{\rm b}\ln(2Up_{\rm b}P_0N\ln N)+0.5772s_{\rm b}]\,. \tag{10}$$

This equation predicts two different regimes for $d\log w(t)/dt$, depending on the values of $s_{\rm b}$ and N. If $s_{\rm b}$ is much smaller than one, and N is only

moderately large (but sufficiently large such that $\mathrm{E}[k] \approx s_\mathrm{b}$), then we can approximate $\ln(1+\mathrm{E}[s])$ with $\mathrm{E}[s]$ and find

$$\frac{d\log w(t)}{dt} \approx s_\mathrm{b}^2[\ln(2Up_\mathrm{b}P_0N\ln N) + 0.5772]\,. \tag{11}$$

In this regime, the speed of adaptation depends logarithmically on N. If on the other hand N is extremely large and s_b is not extremely small, then

$$\frac{d\log w(t)}{dt} \approx s_\mathrm{b}\ln[s_\mathrm{b}\ln(2Up_\mathrm{b}P_0N\ln N)]\,. \tag{12}$$

In this regime, the speed of adaptation depends double-logarithmically on N.

For comparison, I now calculate the speed of adaptation for small N. For small N, clonal interference can be neglected, and therefore the mean beneficial effect corresponds to the mean effect of beneficial mutations that have survived drift. The distribution of these mutations is $g(s) = (s/s_\mathrm{b}^2)\exp(-s/s_\mathrm{b})$ (Rozen et al. 2002; Otto and Jones 2000), and the mean is $\mathrm{E}[s] = 2s_\mathrm{b}$. Using $\ln(1+\mathrm{E}[s]) \approx \mathrm{E}[s]$, we find for small N:

$$\frac{d\log w(t)}{dt} \approx 4s_\mathrm{b}^2NUp_\mathrm{b}P_0\,. \tag{13}$$

To summarize, the speed of adaptation grows linearly in N for small N, and logarithmically or double-logarithmically in N for large N. Interestingly, in the clonal interference regime, growth in the speed of adaptation comes from the fixation of mutations with increasingly larger effects, rather than from the fixation of increasingly more beneficial mutations. Hence, clonal interference slows down the speed of adaptation, but it does not lead to a hard speed limit as long as beneficial mutations of increasingly larger effect are accessible.

Simulation results: I have carried out extensive simulations to test the accuracy of the clonal interference theory. Gerrish and Lenski (1998) mentioned that they found good agreement between theory and simulations, but they did not report any simulation results or the parameter regions they had considered. Orr (2000) reported some simulation results, but his simulations were not in the clonal interference regime [as defined by Eq. (8)].

Figure 3 shows the expected substitution rate as a function of the mutation rate U, both as predicted by Eq. (2) and as found in simulations. Below the optimal mutation rate $U = s_{\mathrm{H}}$ at which the substitution rate assumes its maximum (Orr 2000), agreement between theory and data is good over a wide range of population sizes. Above $U = s_{\mathrm{H}}$, the theory underestimates $\mathrm{E}[k]$. This effect is caused by the accumulation of slightly deleterious mutations in the simulations. The theory assumes that only those beneficial mutations that arise in backgrounds free from deleterious mutations can go to fixation. However, for large U, sequences that carry one or several slightly deleterious mutations become so frequent that it becomes likely that one of them acquires a beneficial mutation of sufficiently large effect to compensate the deleterious background, and goes to fixation. The degree to which the theory underestimates $\mathrm{E}[k]$ increases as s_{H} decreases. In the limit of $s_{\mathrm{H}} = 0$, the theory predicts that $\mathrm{E}[k] = 0$, while simulations show that the true results (with identical s_{b} and s_{d}) are not substantially different from those shown in Fig. 3 (data not shown). Surprisingly, the theory accurately predicts the change in log fitness $d \ln w(t)/dt$ even in the regime of large U, as long as $d \ln w(t)/dt$ is not negative (Fig. 4). [For very high U, Muller's ratchet (Muller 1964; Felsenstein 1974; Haigh 1978; Gordo and Charlesworth 2000) becomes the predominant force in the dynamic of the evolving population, and the change in log fitness can assume negative values.] Apparently, in the

regime of large U, the theory underestimates $\mathrm{E}[k]$ and overestimates $\mathrm{E}[s]$, in such a way that the two effects nearly cancel each other.

In Fig. 5, I show the change in log fitness as a function of the fraction of beneficial mutations p_b. Again, we see excellent agreement between theory and simulation. However, for $p_\mathrm{b} \gtrsim 0.001$ the theory underestimates $\mathrm{E}[k]$ and overestimates $\mathrm{E}[s]$, in such a way that the two effects cancel each other (data not shown).

Finally, in Fig. 6, I show the change in log fitness as a function of the mean effect of new beneficial mutations s_b, while holding the mean effect of new deleterious mutations s_d constant at $s_\mathrm{d} = 0.11$ ($s_\mathrm{H} = 0.05$). As shown in Appendix 1, the theory predicts that both $\mathrm{E}[k]$ and $\mathrm{E}[s]$ should depend linearly on s_b for all parameter values. The change in log fitness should therefore depend quadratically on s_b for $s_\mathrm{b} \lesssim 1$, which means that $d \ln w(t)/dt$ should appear approximately as a straight line with slope 2 in the double-logarithmic plot. We see that the simulation data agree very well with the theory as long as $s_\mathrm{b} \lesssim s_\mathrm{d}$, but start to diverge slowly as s_b grows larger than s_d.

DISCUSSION

Clonal interference is often said to impose a speed limit on adaptation. Here, I have shown that the speed of adaptation, measured as the change in log fitness over time, does not reach a hard limit, but continues to grow even for very large N. This growth is fueled by the discovery of mutations with ever larger beneficial effect in large populations, rather than by an increase in the rate of substitutions.

My results hinge on the assumption that new beneficial mutations are exponentially distributed. If beneficial mutations are distributed such that

large effects are absent, then the rate of adaptation will most likely reach an upper limit for large N. If on the other hand beneficial effects follow a distribution with long tail (such as a power-law or Cauchy distribution), then the speed of adaptation may grow even faster than predicted by Eq. (10) for large N. To date, we do not have a good understanding of the true distribution of beneficial effects in experimental systems. However, an exponential distribution has good theoretical support (Gillespie 1983; Gillespie 1991; Orr 2003), has led to good agreement between theory and experiment in *E. coli* (Rozen et al. 2002), and overall seems to be a reasonable choice.

Arguments for an exponential distribution of new deleterious mutations are not as strong. At the same time, the theory is much less dependent on the particulars of the distribution of deleterious mutations. As long as we have an accurate expression for P_0, and beneficial mutations are unlikely to compensate deleterious mutations, the theory should work. In practice, this means that the theory should work with any distribution that does not produce an excessive amount of slightly deleterious mutations. (Neutral mutations could be dealt with by considering them as a reduction in the overall mutation rate U.)

De Visser et al. (1999) measured the speed of adaptation in *E. coli*, varying both the population size and the mutation rate each over approximately two orders of magnitude. They found that the speed of adaptation did not grow in proportion to increases in population size or mutation rate. In fact, apart from the experiments carried out at the lowest mutation rate, the speed of adaptation changed only very little with population size or mutation rate. These results indicate that the populations with the larger size and higher mutation rates could not benefit from the additional beneficial mutations that must have appeared. The results of de Visser et al. (1999) thus provide

good support for the clonal interference model on a qualitative level. Quantitatively, however, their data seem to disagree with the model analyzed here: Fig. 2A of de Visser et al. (1999) suggests that the speed of adaptation runs quickly into a hard limit, whereas the model predicts that the speed should continue to grow logarithmically, at least with respect to population size.

There are two reasons that may have caused this discrepancy. First, de Visser et al. (1999) plotted the speed of adaptation versus the relative mutation-supply rate (which is the product of population size and relative mutation rate). Such a plot is problematic, because the mutation-supply rate does not uniquely specify the speed of adaptation in the clonal interference model. [In order for the mutation-supply rate to uniquely specify the speed of adaptation, population size and mutation rate would have to enter the equations always as a product, which is not the case. In Eq. (10), for example, the term P_0 depends on U but not on N, while the term $\ln N$ does not depend on U.] The model predicts that the speed of adaptation should increase with increasing N, whereas it should reach a maximum and then decrease with increasing U. Thus, a plot of the speed of adaptation versus population size (at fixed U) is inherently more informative than a plot of the speed of adaptation versus mutation rate (at fixed N). In the latter case, a decline in the increase of the speed of adaptation may also indicate that the mutation rate approaches the optimal mutation rate $U = s_{\mathrm{H}}$. Since de Visser et al. (1999) studied only two different population sizes (and three mutation rates), it is not possible to replot a subset of their data versus population size at fixed U and obtain a quantitative comparison to the model.

Second, the speed of adaptation at a high mutation-supply rate may have been reduced (thus giving the impression of a hard speed limit) in part because the populations began to run out of beneficial mutations. De Visser

et al. propagated the populations for 1000 generations, and determined the speed of adaptation from the total fitness increase over these 1000 generations. In particular for the large population size, it seems that adaptation slowed down considerably after 500 generations (de Visser et al. 1999, Fig. 1B). [Note, however, that this argument does not invalidate the overall conclusion of de Visser et al. (1999). The fitness increase after 200 generations in their Fig. 1 shows strong dependence on the mutation rate for the small population size, and weak dependence on the mutation rate for the large population size, which agrees very well with the predictions of the clonal interference model.] A related explanation for the apparent slowdown in the speed of adaptation at large population size is that the large populations may have found mutations of large beneficial effect earlier in the experiments than the small populations, as predicted by the clonal interference model.

The clonal interference model assumes an infinite supply of beneficial mutations, and this assumption is of course unrealistic. Nevertheless, we can expect good agreement between model and experiment if the experiment is restricted to a relatively short number of generations, or if only the effect of the first fixed mutation is measured. An experiment of the latter kind was carried out by Rozen et al. (2002), who found that the measured distribution of beneficial effects in *E. coli* was in good agreement with the distribution as predicted by the clonal interference model.

The data of Rozen et al. (2002) also allows us to estimate the onset of clonal interference in *E. coli*. By fitting the theoretical prediction for the distribution of beneficial effects to their data, Rozen et al. (2002) derived estimates for the mean beneficial effect of new mutations s_b and for the beneficial mutation rate Up_b. They found $s_\mathrm{b} = 0.024$ and $Up_\mathrm{b} = 5.9 \times 10^{-8}$. Having an estimate for the beneficial mutation rate, we can use Eq. (8) to estimate

the population size in these *E. coli* populations at which clonal interference becomes important. Since we do not have a good estimate for the mutational load in these populations, we set $P_0 = 1$, which means that we neglect the effect of deleterious mutations. [See Orr (2000) for a discussion of this problem and its implications for the estimates of s_b and Up_b.] As a consequence, we most likely underestimate the population size at which clonal interference becomes important. Further, instead of the factor 2 in front of Up_b, we use 0.6. This factor takes into account that the *E. coli* populations fluctuate in size under standard laboratory conditions, see Rozen et al. (2002), p. 1044. Thus, we use for our estimate $N \ln N > 1/(0.6 \times 5.9 \times 10^{-8})$. This condition simplifies to $N > 2 \times 10^6$. [Using a less accurate method based on only the expected substitution rate and mean beneficial effect of fixed mutations, Gerrish and Lenski (1998) had earlier derived an estimate of $Up_\mathrm{b} = 2 \times 10^{-9}$, which leads to $N > 4.7 \times 10^7$.] Rozen et al. (2002) carried out their experiments at an effective population size of $N = 3.3 \times 10^7$, which means that clonal interference probably had an effect on their results. This reasoning is consistent with the observation that the mean beneficial effect of fixed mutations is clearly larger than $2s_\mathrm{b}$ in their data (Rozen et al. 2002, Fig. 3).

Clonal interference has not only been studied in *E. coli*, but also in vesicular stomatitis virus (VSV). Following de Visser et al. (1999), Miralles et al. (1999) fitted a linear and a hyperbolic model to the rate of fitness change in VSV as a function of population size, and found that the hyperbolic model provided the better fit. However, their data does not plateau at high N, and visual inspection of their Fig. 1 suggests that a logarithmic model might fit their data as well. On the other hand, the multiplicity of infection was changed alongside with the population size in these experiments, so that a slow-down in the speed of adaptation could also be due to increased virus-virus interactions

within cells, rather than clonal interference (Wilke and Novella 2003).

My simulations have shown that the theory of clonal interference works well for small to moderate mutation rates, but fails at high mutation rates, when Muller's ratchet becomes important. Another effect at high mutation rates that is neglected in the theory (but was also absent from the simulations) is the evolution of mutational robustness: If the distribution of deleterious mutations itself can change, then at a high mutation rate there is a selective pressure to minimize the mutational load of the population (van Nimwegen et al. 1999; Wilke et al. 2001; Wilke and Adami 2003). This effect will increase the mean fitness of the population, and will also increase the potential for further adaptation by increasing P_0. In general, at a high mutation rate we have to consider that a mutation will be combined with additional mutations on the way to fixation. Therefore, we cannot simply assume that a mutation of beneficial effect s has probability of fixation $2s$, but have to use fairly complicated mathematical tools (such as multi-type branching processes) to calculate the fixation probability (Barton 1995; Johnson and Barton 2002; Wilke 2003; Iwasa et al. 2004). As a consequence, it is unlikely that we will ever have a simple closed-form expression for the speed of adaptation at a high mutation rate.

A second regime in which the theory—not surprisingly—breaks down is when s_b exceeds s_d. In this regime, we cannot simply neglect all beneficial mutations that do not arise on genetic backgrounds free of deleterious mutations. Johnson and Barton (2002) have recently studied this situation, but not in the clonal interference regime. In the clonal interference regime, there are two opposing effects to be considered: On the one hand, the total number of competing mutations should increase, since now beneficial mutations on deleterious backgrounds compete for fixation as well. On the other hand, the

total fitness effect of many of the mutations that are competing for fixation is smaller than what we expect from the distribution of new beneficial mutations, because the beneficial effects are reduced by deleterious backgrounds. Without a detailed analysis, it is unclear which of the two effects is more important. However, if the results from the present study are an indication, in the clonal interference regime the number of competing mutations will be less important than the distribution of their beneficial effects, which means that the present theory should overestimate the rate of adaptation for $s_\mathrm{b} > s_\mathrm{d}$. Indeed, I observed exactly this behavior in my simulations (Fig. 6).

Whether the assumption $s_\mathrm{b} < s_\mathrm{d}$ is reasonable is not yet resolved. (Likely, the answer to this question will also depend on the particular species under study and on the concrete selection regime.) Several authors found that deleterious mutations were frequent but had a very small effect (Mukai et al. 1972; Ohnishi 1977; Kibota and Lynch 1996; Shabalina and Kondrashov 1997; Elena and Moya 1999), while others found that deleterious mutations were less frequent but of larger effect (Keightley 1996; Fernández and López-Fanjul 1996; García-Dorado 1997; Keightley and Caballero 1997). If the first set of results is representative, then beneficial mutations may indeed have on average a larger effect on fitness than deleterious mutations. While it is reasonably straightforward to study the distribution of deleterious mutations in mutation-accumulation experiments, it is much harder to measure the distribution of new beneficial mutations (as opposed to the distribution of fixed beneficial mutations, which is skewed towards mutations of large effect). However, evidence from experimental evolution with viruses shows that in some cases, beneficial mutations must have very large effects: Wichman et al. (1999) found several-thousand-fold increase in population growth after ten days of selection in phage ΦX174, and Novella et al. (1995) found fitness increases by a factor of 10 or more

within five to ten generations in vesicular stomatitis virus. In both cases, the observed fitness increase within a very short time frame can only be explained by a large supply of beneficial mutations of large effect.

To summarize, in this contribution I have found the following novel conclusions:

1. The expected rate of adaptation approaches the mean beneficial effect of new mutations for large N.

2. The mean beneficial effect of fixed mutations grows logarithmically in N for large N.

3. Clonal interference effects become important if $N \ln N$ is larger than $1/(2Up_{\mathrm{b}}P_0)$.

4. The speed of adaptation grows logarithmically in N for moderately large N, and double-logarithmically for extremely large N.

5. For large N, the speed of adaptation is limited by the distribution of beneficial effects of new mutations rather than by the supply rate of new mutations.

ACKNOWLEDGMENTS

I would like to thank H.A. Orr for clarifications regarding his work, and S.F. Elena, R.E. Lenski, S. P. Otto, J.A.G.M. de Visser, and two anonymous reviewers for helpful comments on earlier versions of this article. This work was supported by the NSF under contract No. DEB-9981397.

References

Barton, N. H., 1995 Linkage and the limits to natural selection. Genetics **140:** 821-841.

Bull, J. J., M. R. Badgett and H. A. Wichman, 2000 Big-benefit mutations in a bacteriophage inhibited with heat. Mol. Biol. Evol. **17:** 942-950.

Campos, P. R. A., 2003 Fixation of beneficial mutations in the presence of epistatic interactions. Bull. Math. Biol., in press.

Crow, J. F., and M. Kimura, 1965 Evolution in sexual and asexual populations. Am. Nat. **99:** 439-450.

de Visser, J. A. G. M., C. W. Zeyl, P. J. Gerrish, J. L. Blanchard and R. E. Lenski, 1999 Diminishing returns from mutation supply rate in asexual populations. Science **283:** 404-406.

Elena, S. F., and A. Moya, 1999 Rate of deleterious mutation and the distribution of its effects on fitness in vesicular stomatitis virus. J. Evol. Biol. **12:** 1078-1088.

Felsenstein, J., 1974 The evolutionary advantage of recombination. Genetics **78:** 737-756.

Fernández, J., and C. López-Fanjul, 1996 Spontaneous mutational variances and covariances for fitness-related traits in *Drosophila melanogaster*. Genetics **143:** 829-837.

García-Dorado, A., 1997 The rate and effects distribution of viable mutation in *Drosophila*: Minimum distance estimation. Evolution **51:** 1130-1139.

Gerrish, P., 2001 The rhythm of microbial adaptation. Nature **413:** 299-302.

Gerrish, P. J., and R. E. Lenski, 1998 The fate of competing beneficial mutations in an asexual population. Genetica **102/103:** 127-144.

Gillespie, J. H., 1983 A simple stochastic gene substitution model. Theor. Popul. Biol. **23:** 202-215.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution.* Oxford University Press, New York.

Gordo, I., and B. Charlesworth, 2000 The degeneration of asexual haploid populations and the speed of Muller's ratchet. Genetics **154:** 1379-1387.

Haigh, J., 1978 The accumulation of deleterious genes in a population – Muller's ratchet. Theor. Popul. Biol. **14:** 251-267.

Hill, W. G., and A. Robertson, 1966 The effect of linkage on limits to artificial selection. Genet. Res. **8:** 269-294.

Imhof, M., and C. Schlötterer, 2001 Fitness effects of advantageous mutations in evolving *Escherichia coli* populations. Proc. Natl. Acad. Sci. USA **98:** 1113-1117.

Iwasa, Y., F. Michor and M. A. Nowak, 2004 Evolutionary dynamics of invasion and escape. J. theor. Biol. **226:** 205-214.

Johnson, T., and N. H. Barton, 2002 The effect of deleterious alleles on adaptation in asexual populations. Genetics **162:** 395-411.

Keightley, P. D. 1996 Nature of deleterious mutation load in *Drosophila*. Genetics **144:** 1993-1999.

Keightley, P. D., and A. Caballero, 1997 Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **94:** 3823-3827.

Kibota, T. T., and M. Lynch, 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. Nature **381:** 694-696.

Kichler Holder, K., and J. J. Bull, 2001 Profiles of adaptation in two similar viruses. Genetics **159:** 1393-1404.

Kim, Y., and W. Stephan, 2003 Selective sweeps in the presence of interference among partially linked loci. Genetics **164:** 389-398.

Kimura, M., and T. Ohta, 1971 *Theoretical Aspects of Population Genetics*. Princeton University Press, Princeton, N.J.

Kirby, D. A., and W. Stephan, 1996 Multi-locus selection and the structure of variation at the *white* gene of *Drosophila melanogaster*. Genetics **144:** 635-645.

McVean, G. A. T., and B. Charlesworth, 2000 The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics **155:** 929-944.

Miralles, R., P. J. Gerrish, A. Moya and S. F. Elena, 1999 Clonal interference and the evolution of RNA viruses. Science **285:** 1745-1747.

Miralles, R., A. Moya and S. F. Elena, 2000 Diminishing returns of population size in the rate of RNA virus adaptation. J. Virol. **74:** 3566-3571.

Mukai, T., S. I. Chigusa, L. E. Mettler and J. F. Crow, 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. Genetics **72:** 335-355.

Muller, H. J., 1964 The relation of recombination to mutational advance. Mutation Research **1:** 2-9.

Novella, I. S., E. A. Duarte, S. F. Elena, A. Moya, E. Domingo and J. J. Holland, 1995 Exponential increases of RNA virus fitness during large

population transmissions. Proc. Natl. Acad. Sci. USA **92:** 5841-5844.

Ohnishi, O., 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster*. II. Homozygous effect of polygenic mutations. Genetics **87:** 529-545.

Orr, H. A., 2000 The rate of adaptation in asexuals. Genetics **155:** 961-968.

Orr, H. A., 2003 The distribution of fitness effects among beneficial mutations. Genetics **163:** 1519-1526.

Orr, H. A., and Y. Kim, 1998 An adaptive hypothesis for the evolution of the Y chromosome. Genetics **150:** 1693-1698.

Otto, S. P., and C. D. Jones, 2000 Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. Genetics **156:** 2093-2107.

Rozen, D. E., J. A. G. M. de Visser and P. J. Gerrish, 2002 Fitness effects of fixed beneficial mutations in microbial populations. Curr. Biol. **12:** 1040-1045.

Shabalina, S. A., and L. Y. Y. A. S. Kondrashov, 1997 Rapid decline of fitness in panmictic populations of *Drosophila melanogaster* maintained under relaxed natural selection. Proc. Natl. Acad. Sci. USA **94:** 13034-13039.

Stephan, W., 1995 Perturbation analysis of a 2-locus model with directional selection and recombination. J. Math. Biol. **34:** 95-109.

van Nimwegen, E., J. P. Crutchfield and M. Huynen, 1999 Neutral evolution of mutational robustness. Proc. Natl. Acad. Sci. USA **96:** 9716-9720.

Wichman, H. A., M. R. Badgett, L. A. Scott, C. M. Boulianne and J. J. Bull, 1999 Different trajectories of parallel evolution during viral adaptation. Science **285:** 422-424.

Wilke, C. O., 2003 Probability of fixation of an advantageous mutant in a viral quasispecies. Genetics **163:** 467-474.

Wilke, C. O., and C. Adami, 2003 Evolution of mutational robustness. Mutation Research **522:** 3-11.

Wilke, C. O., and I. S. Novella, 2003 Phenotypic mixing and hiding may contribute to memory in viral quasispecies. BMC Microbiology **3:** 11.

Wilke, C. O., J. L. Wang, C. Ofria, R. E. Lenski and C. Adami, 2001 Evolution of digital organisms at high mutation rate leads to survival of the flattest. Nature **412:** 331-333.

APPENDIX 1

Expected substitution rate and mean beneficial effect: First, we notice that E[k] depends linearly on the mean effect of beneficial mutations: Substituting x for s/s_b in Eq. (2), and writing $A = 2Up_\mathrm{b}P_0N\ln N$, we find

$$\mathrm{E}[k] = \frac{s_\mathrm{b}A}{\ln N}\int_0^\infty x\exp[-A(1+1/x)e^{-x}-x]dx\,. \tag{14}$$

Therefore, the shape of E[k] as a function of N or U is independent of the value of s_b.

We are interested in an asymptotic expansion of the integral in Eq. (14) for large N. For the asymptotic expansion to work, N needs to be so large that A is large. Clearly, for any given $Up_\mathrm{b}P_0$, we can always choose N sufficiently large such that A is large. Because of the exponential factor in the integrand, the main contribution to the integral comes from values of x for which $A(1+1/x)e^{-x}+x$ is small. Since the first term decays exponentially with x while the second term grows linearly, in general the main contribution to the integral will come from small x. However, for large A, the first term becomes small only when x is substantially larger than one. In this regime, we can neglect the term $1/x$, and the integral in Eq. (14) is then identical to the integral $J_n(A)$ defined in Appendix 2 with $n = 1$. Using the expression for $J_1(A)$ given in Eq. (23), we find

$$\mathrm{E}[k] \approx \frac{s_\mathrm{b}}{\ln N}[\ln(2Up_\mathrm{b}P_0N\ln N)+\gamma]\,. \tag{15}$$

In the limit of very large N, we obtain the even simpler expression E[k] $\approx s_\mathrm{b}$.

Using a reasoning similar to that for the expected substitution rate, we find that the expected beneficial effect [as given in Eq. (5)] also depends linearly on s_b, and simplifies for large N to E[s] $= s_\mathrm{b}J_2(A)/J_1(A)$ (again

with $A = 2Up_bP_0N\ln N$). Using the expressions given in Eqs. (22) and (23), we find

$$\mathrm{E}[s] \approx s_b[\ln(2Up_bP_0N\ln N) + \gamma] + \frac{s_b\pi^2/6}{\ln(2Up_bP_0N\ln N) + \gamma}\,. \qquad (16)$$

In the limit of very large N, the second term disappears, and we end up with

$$\mathrm{E}[s] \approx s_b[\ln(2Up_bP_0N\ln N) + \gamma]\,. \qquad (17)$$

Estimating the onset of clonal interference: We can derive an estimate of the parameter region in which clonal interference becomes important by calculating the point at which the approximation for E[k] for small N [Eq. (1)] comes the closest to the approximation for E[k] for large N, Eq. (3). Since the shape of E[k] is not influenced by s_b (see above), we can set $s_b = 1$ for this calculation. Further, we write $C = 2Up_bP_0$. Now, we have to find the minimum of the function

$$g(C,N) = CN - [\ln(CN\ln N) + \gamma]/\ln N\,. \qquad (18)$$

We find $\partial g(C,N)/\partial C = N - 1/(C\ln N)$, which leads to the condition

$$N\ln N > \frac{1}{C} \qquad (19)$$

for the onset of clonal interference. (Differentiating with respect to N yields approximately the same condition). This condition cannot be solved for N in a closed-form expression, but is easy to evaluate numerically.

APPENDIX 2

Integrals: For the asymptotic expansion, we have to solve integrals of the form

$$J_n(A) = \int_0^\infty x^n \exp(-Ae^{-x} - x)dx\,, \qquad (20)$$

in particular for the cases $n = 1$ and $n = 2$. After substituting $z = Ae^{-x}$, we obtain

$$J_n(A) = \frac{1}{A}\int_0^A (\ln A - \ln z)^k e^{-z} dz$$
$$= \frac{1}{A}\sum_{k=0}^{n} \binom{n}{k} (\ln A)^{n-k}(-1)^k \int_0^A (\ln z)^k e^{-z} dz\,. \qquad (21)$$

The main contribution to the remaining integral comes from small z, while A is large in the cases considered here. Therefore, we can replace the upper limit of integration with ∞. For the three relevant cases $k = 0$, $k = 1$, and $k = 2$, the integrals are $\int_0^\infty e^{-z} dz = 1$, $\int_0^\infty \ln z\, e^{-z} dz = -\gamma$, $\int_0^\infty (\ln z)^2 e^{-z} dz = \gamma^2 + \pi^2/6$, where $\gamma \approx 0.5772$ is the Euler constant. Thus, we find approximately

$$J_1(A) = (\ln A + \gamma)/A\,, \qquad (22)$$
$$J_2(A) = [(\ln A + \gamma)^2 + \pi^2/6]/A\,. \qquad (23)$$

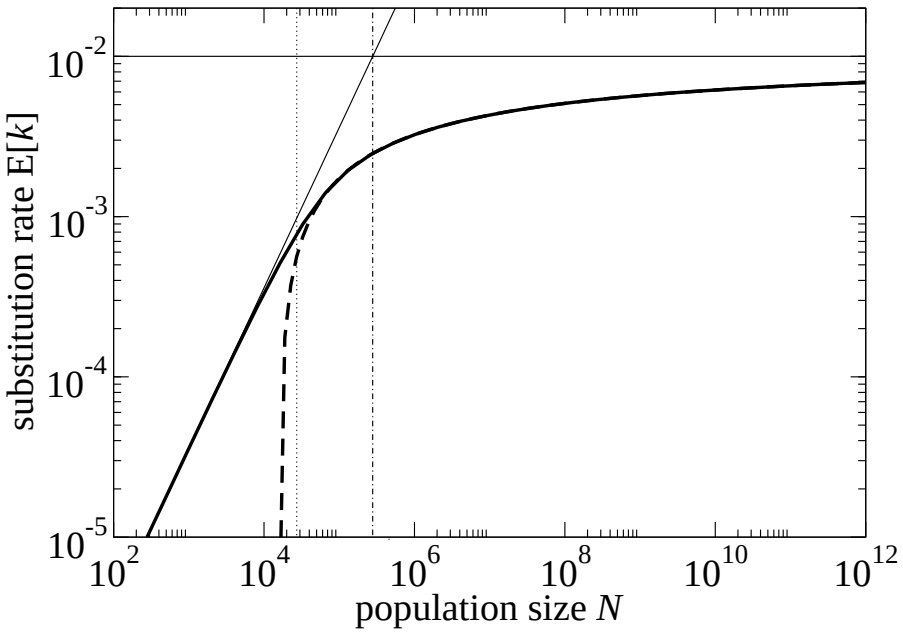

Figure 1: Expected substitution rate E[k] versus population size N ($U = 0.04$, $p_\mathrm{b} = 0.0001$, $s_\mathrm{b} = 0.01$, $s_\mathrm{H} = 0.05$). The thick solid line stems from exact numerical evaluation of Eq. (2), and the thick dashed line corresponds to approximation Eq. (3). The thin solid lines correspond to the approximations for small and large N, Eqs. (1) and (4). The dash-dotted line indicates the onset of clonal interference according to Eq. (7), and the dotted line indicates the same according to Eq. (8).

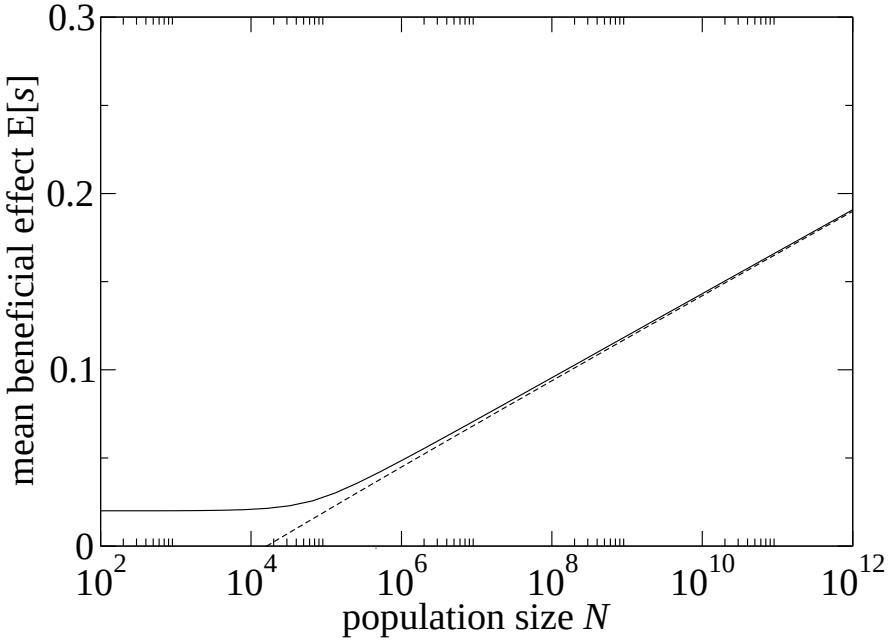

Figure 2: Mean beneficial effect of fixed mutations E[s] versus population size N ($U = 0.04$, $p_\mathrm{b} = 0.0001$, $s_\mathrm{b} = 0.01$, $s_\mathrm{H} = 0.05$). The solid line stems from exact numerical evaluation of Eq. (5), and the dashed line corresponds to approximation Eq. (6).

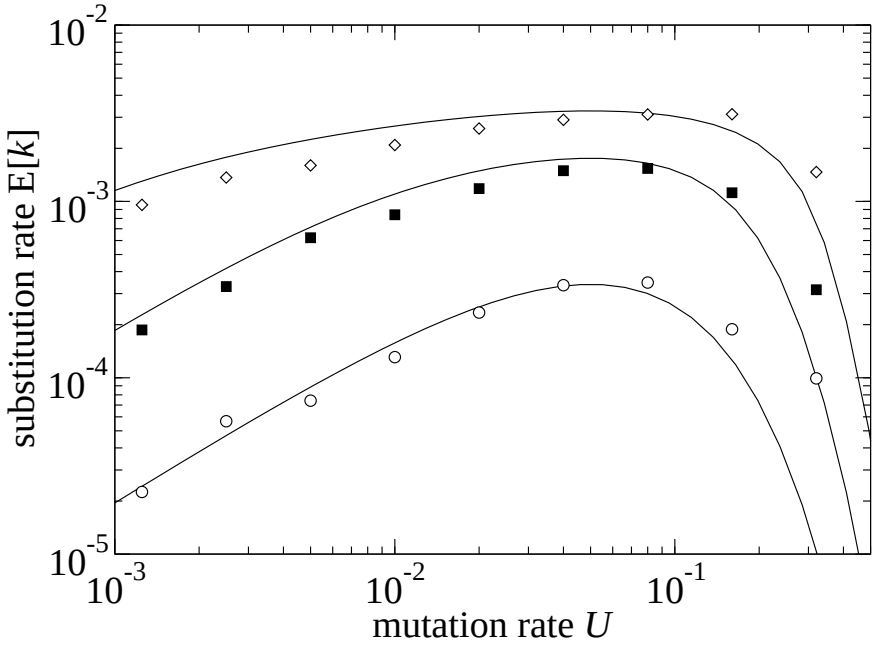

Figure 3: Expected substitution rate E[k] versus mutation rate U ($p_{\rm b}$ = 0.0001, $s_{\rm b}$ = 0.01, $s_{\rm H}$ = 0.05). Population sizes are (from bottom to top) $N = 10^4$, $N = 10^5$, $N = 10^6$. Solid lines indicate the theoretical prediction Eq. (2), and points are simulation results.

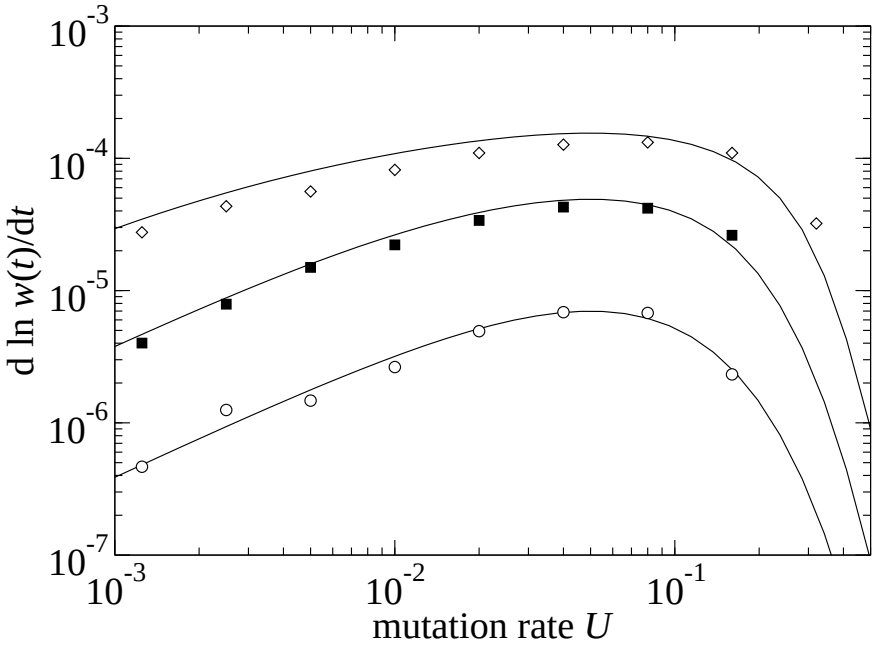

Figure 4: Change in log fitness $d \ln w(t)/dt$ versus mutation rate U ($p_{\mathrm{b}} = 0.0001$, $s_{\mathrm{b}} = 0.01$, $s_{\mathrm{H}} = 0.05$). Population sizes are (from bottom to top) $N = 10^4$, $N = 10^5$, $N = 10^6$. Solid lines indicate the theoretical prediction $\mathrm{E}[k] \ln(1 + \mathrm{E}[s])$, and points are simulation results. For $U = 0.4$, Muller's ratchet led to a negative $d \ln w(t)/dt$ in the populations of size $N = 10^4$ and $N = 10^5$. The corresponding two data points are therefore missing from this figure.

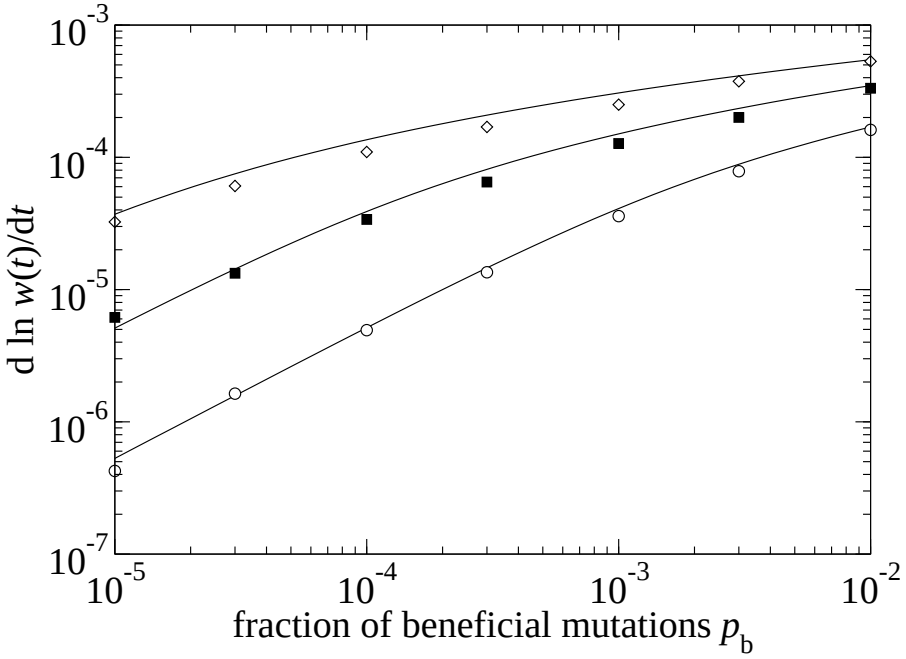

Figure 5: Change in log fitness $d \ln w(t)/dt$ versus fraction of beneficial mutations p_b ($U = 0.02$, $s_\mathrm{b} = 0.01$, $s_\mathrm{H} = 0.05$). Population sizes are (from bottom to top) $N = 10^4$, $N = 10^5$, $N = 10^6$. Solid lines indicate the theoretical prediction $\mathrm{E}[k] \ln(1 + \mathrm{E}[s])$, and points are simulation results.

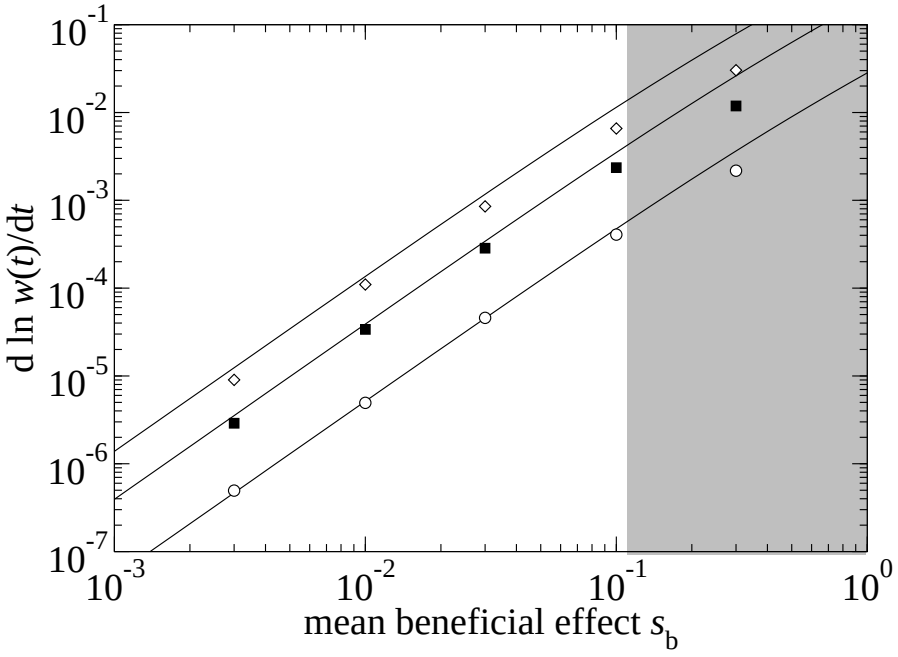

Figure 6: Change in log fitness $d \ln w(t)/dt$ versus mean effect of new beneficial mutations s_{b} ($p_{\mathrm{b}} = 0.0001$, $U = 0.02$, $s_{\mathrm{H}} = 0.05$). Population sizes are (from bottom to top) $N = 10^4$, $N = 10^5$, $N = 10^6$. Solid lines indicate the theoretical prediction $\mathrm{E}[k] \ln(1 + \mathrm{E}[s])$, and points are simulation results. In the shaded region, the mean effect of new beneficial mutations s_{b} exceeds the mean effect of new deleterious mutations s_{d}.